



The where and the when of loanword usage in dialectal varieties

Karliën Franco, Dirk Geeraerts & Dirk Speelman



RU Quantitative Lexicology and Variational Linguistics

ONSCHULDIG

'innocent'



onschuldig

kuis

zebedeus

daar zit geen kwaad in

snulletje

onnozel

simpel

18 different words
(Lim. & Brab.)

BANGERIK

'coward'



bange

bangerd

angstige

bange floets

schouwe

bange pezerik

bang schijthuis

held op sokken

angstpiemel

bangboks

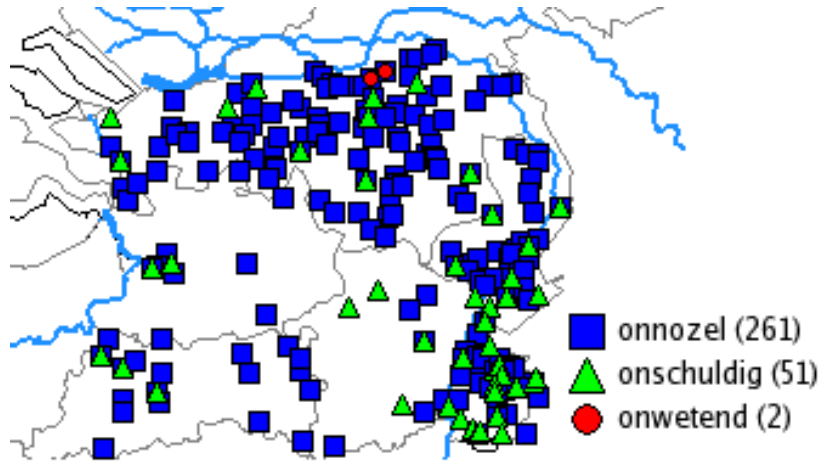
schrikkepee

angstschijter

100 different words
(Lim. & Brab.)

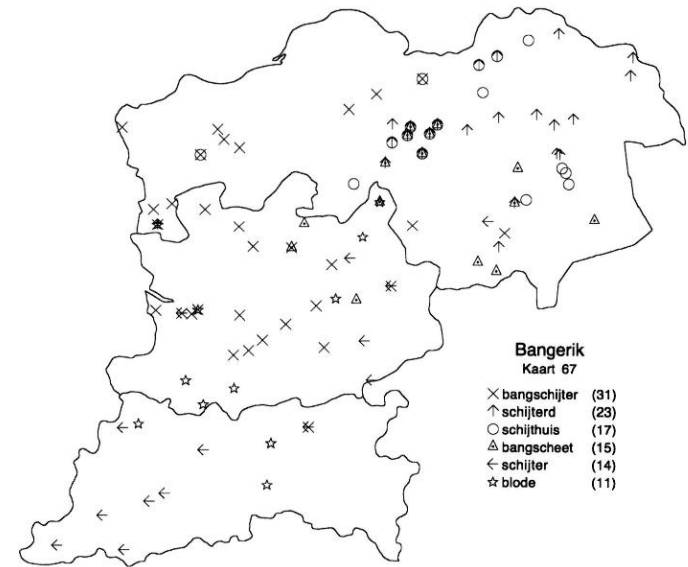


ONSchULDIG 'innocent'



two variants that occur
nearly everywhere

BANGERIK 'coward'



small geographical
areas



can naming strategies provide further insight?

there is structure in naming strategies

- names for birds reflect how well-known a bird is
- similar patterns occur for names of clothes
- plant names are often based on the shape or color of the plant
- useful plants (i.e. edible plants or plants with medicinal applications) show less lexical variation

→ naming strategies show how language users structure their daily environment

(Geeraerts, Grondelaers & Bakema 1994, Swanenberg 2000, Brok 1993, Franco et al. 2017)



when?

borrowing as a naming strategy

- necessary and luxury loans

cheerleader vs. freak (zonderling)

- the success of a loanwords differs per semantic field

Latin: a.o. **christianity** e.g. *evangelie, kardinaal, klooster*

military e.g. *defensie, pijl*

French a.o. **ME courts** e.g. *baldakijn, buffet, kasteel*

administration e.g. *parket, parlement*

clothing e.g. *mannequin, jupon, bretel*

- diachronic differences

(Van der Sijs 1996, Zenner, Speelman & Geeraerts 2012)

where?

geographical differences in loanword usage

- more intense language contact with French in Flanders than in the Netherlands apparent from the higher number of French loans in Spoken Belgian Dutch

vb. *camion, kravat, gazet*

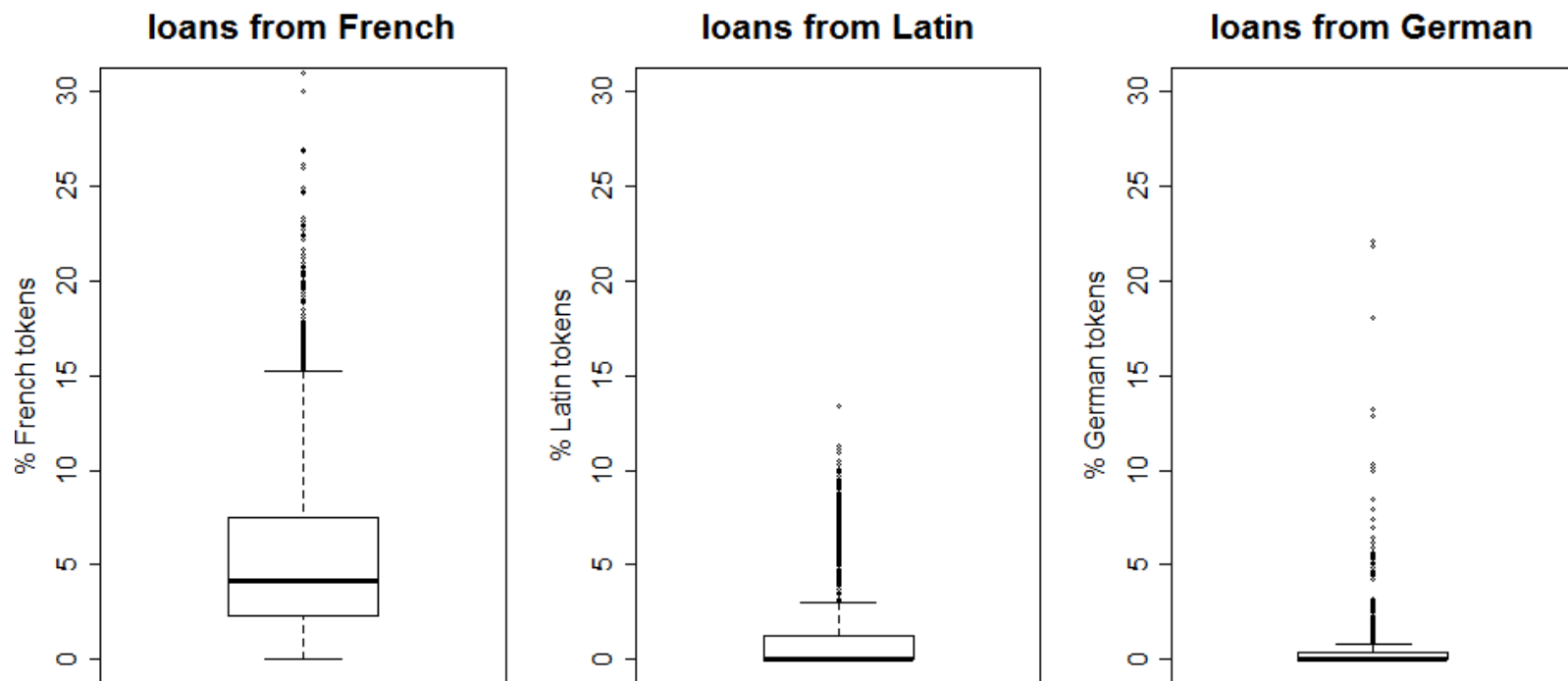
N.B. purism

- more language contact near language borders
but state border can evolve into a dialect border

(Weijnen & Van Coetsem 1957, Giesbers 2008, Van der Sijs 1996)



can we find structure in the usage of loanwords?



geographical structure? semantic structure?

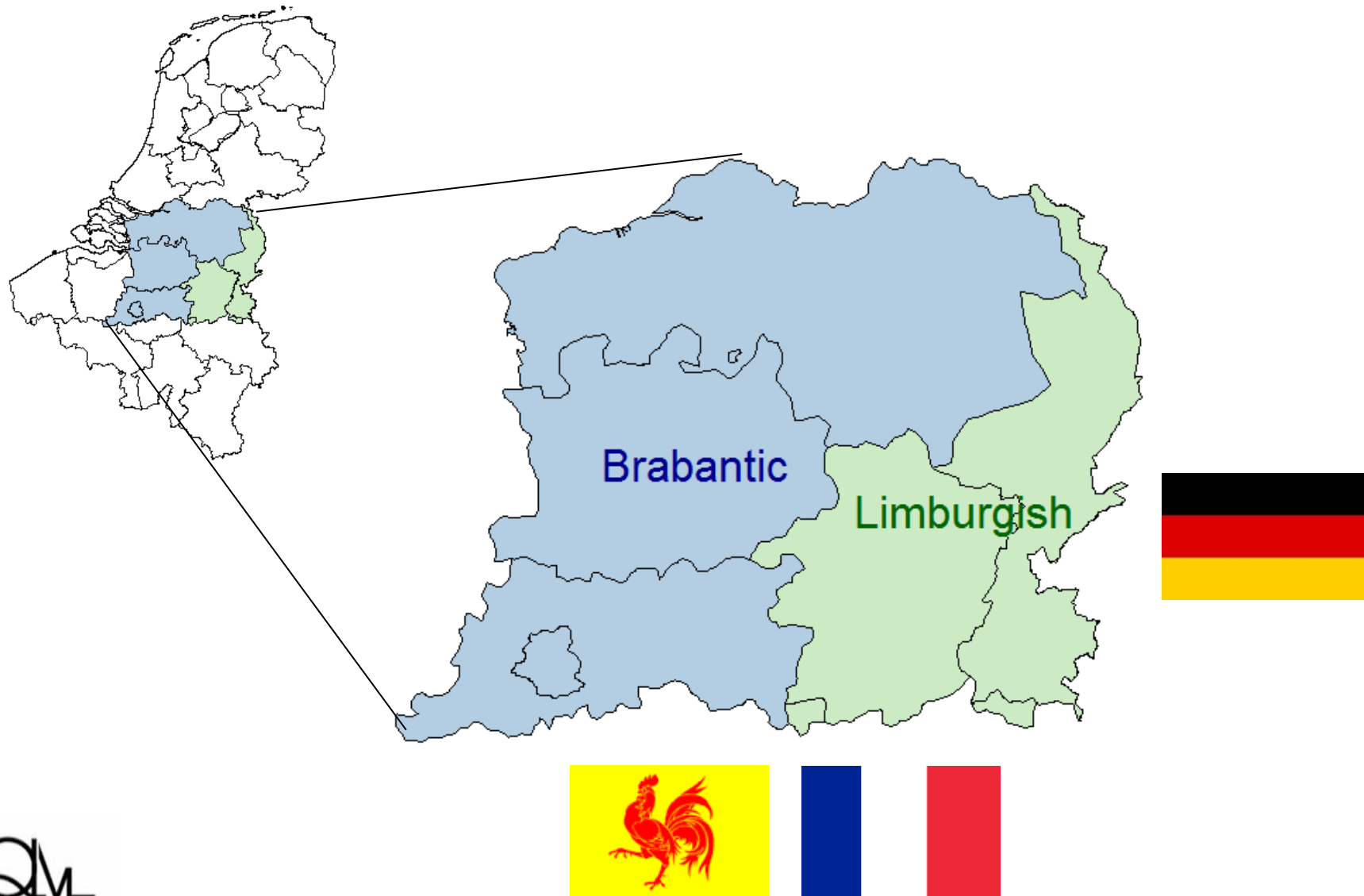
we expect...

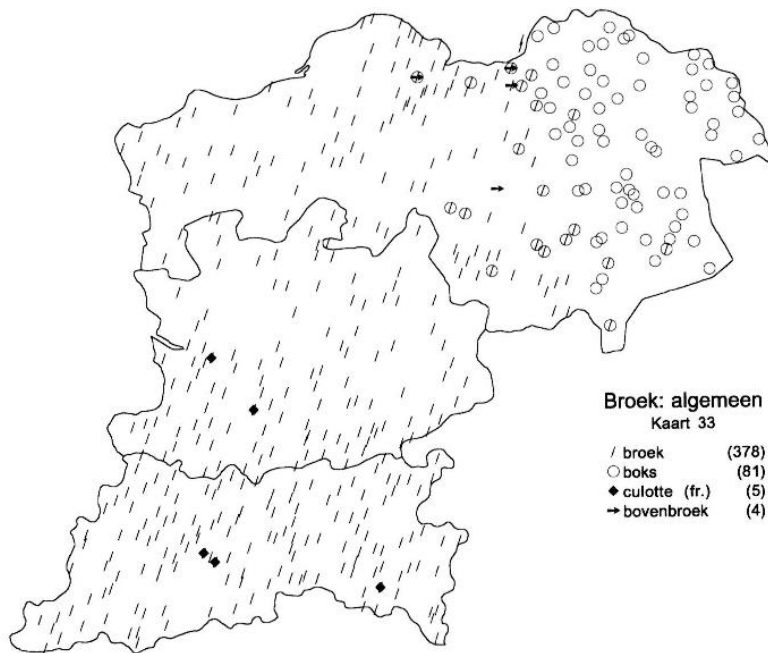
- geographical patterns
 - more French in Flanders than in the Netherlands
 - more German close to the border with Germany
 - no geographical structure in the usage of Latin
- differences between semantic fields
 - more French for clothing terms and (mostly in Flanders) for concepts relating to society and education
 - more Latin for concepts concerning church & religion

the dialects of Dutch



the dialects of Dutch





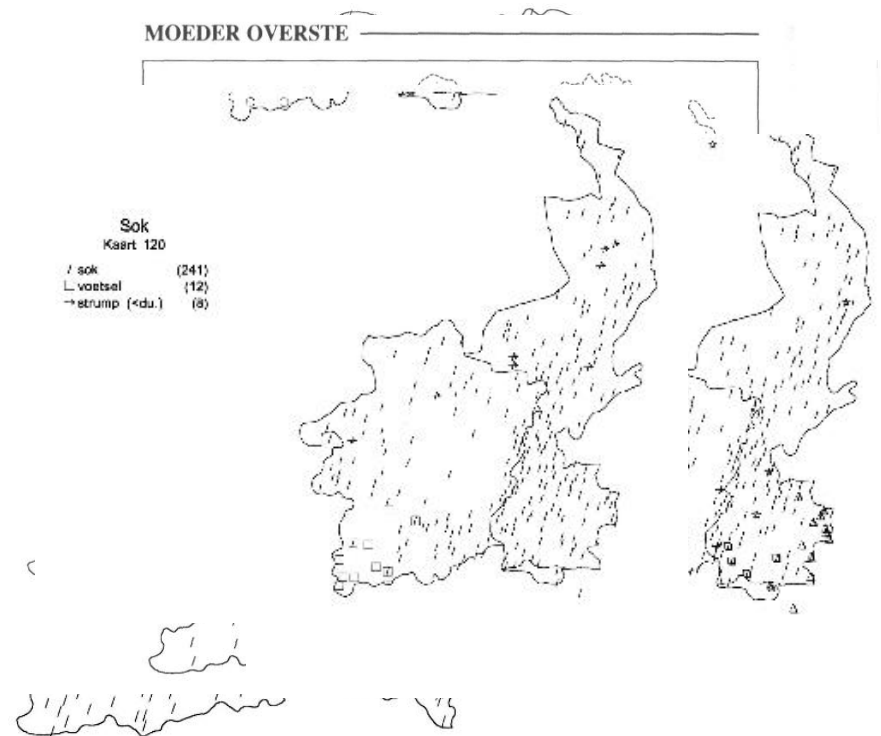
traditional dialectology

interpretation **per concept**

detailed analysis

detailed **etymology**

qualitative



"big data dialectology"

aggregation: many concepts at once

general patterns on the basis of analyses and etymologies that have already been conducted

quantitative

(Generalized Additive Modelling)

in practice

concept	variant	location	...
damesmantel 'coat for women'	caban (fr.)	Tervuren	...
overjas 'overcoat'	frak	Leopoldsburg	...
...
vrolijk 'cheerful'	spass (du.) haan	Simpelveld	...
vrolijk 'cheerful'	opgewekt	Venlo	...
...
heilige hostie 'sacred host'	hostie (lat.)	Bocholt	...
heilige hostie 'sacred host'	Ons Lieve Heer	Neerpelt	...
...



in practice

concept	variant	location	...
damesmantel 'coat for women'	caban (fr.)	Tervuren	...
overjas 'overcoat'	frak	Leopoldsburg	...
...
vrolijk 'cheerful'	spass (du.) haan	Simpelveld	...
vrolijk 'cheerful'	opgewekt	Venlo	...
...
heilige hostie 'sacred host'	hostie (lat.)	Bocholt	...
heilige hostie 'sacred host'	Ons Lieve Heer	Neerpelt	...
...



how 'big' is the big data?

543 659 words (tokens)

43 828 different words (types)

2 338 concepts

637 locations

221 368 Brabantian tokens

322 291 Limburgish tokens

29 458 French tokens

10 171 Latin tokens

2 635 German tokens

analyze the proportion of
French/Latin/German variants per
location

e.g. largest proportion of French occurs in Vorsen (over 30% of all tokens)

- 'combinaison' (ONDERJURK) vs. 'onderrok' & 'onderkleed'
- 'bijou' (JUWEEL) vs. 'juweel' & 'edelsteen'
- 'pardessus' (OVERJAS) vs. 'overjas'



Generalized Additive Modelling (GAM)

- extension of GLMs, which allows for more complex relationships between predictors and response (wiggleness)
- one model per source language (French, Latin, German)
- basic model:

proportion of loanwords per location ~

semantic field +

smooth term for lon*lat by semantic field +

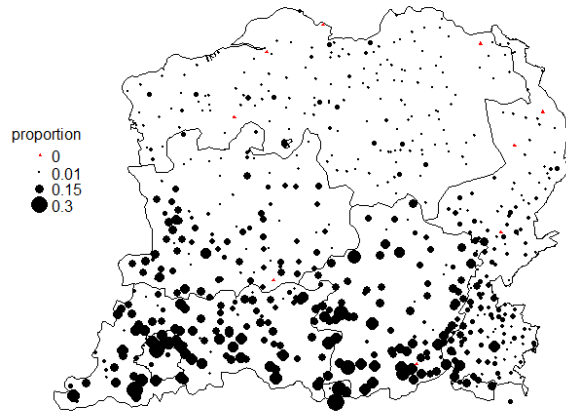
random intercept for location (NS for Latin)

(Crawley 2007, Faraway 2006, Wood 2006, Wieling 2012, Zuur et al. 2009)

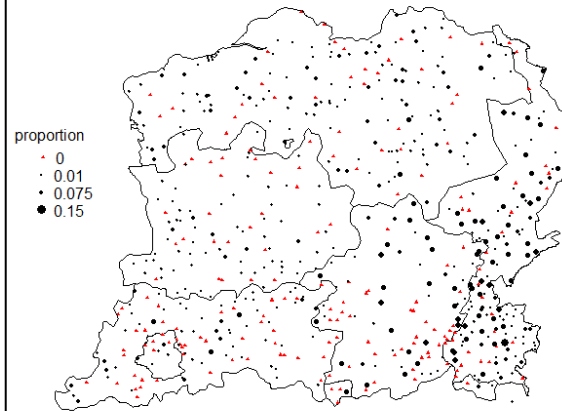


the general picture

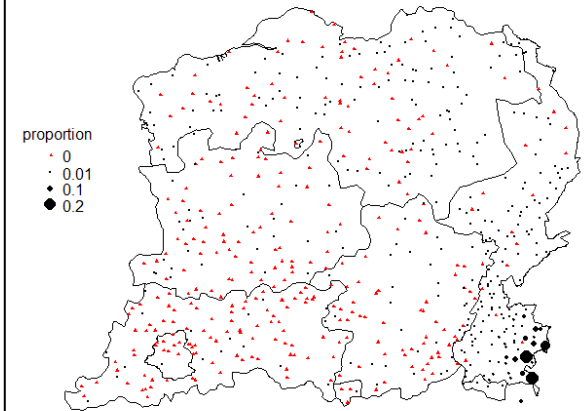
Proportion of French tokens per location



Proportion of Latin tokens per location



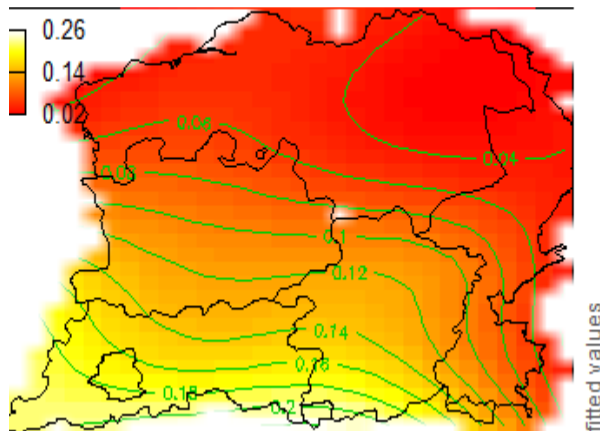
Proportion of German tokens per location



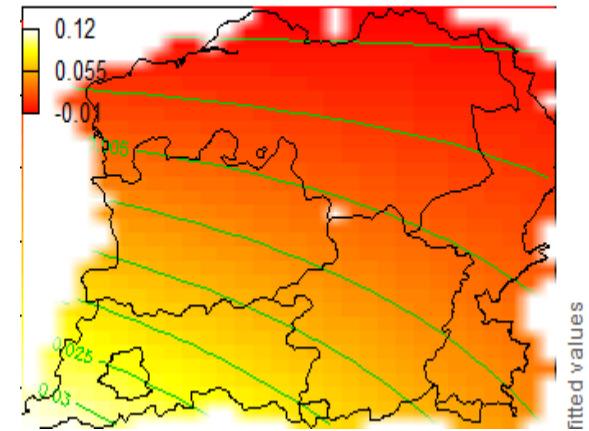
deviance explained: 89.6%

semantic patterns: French

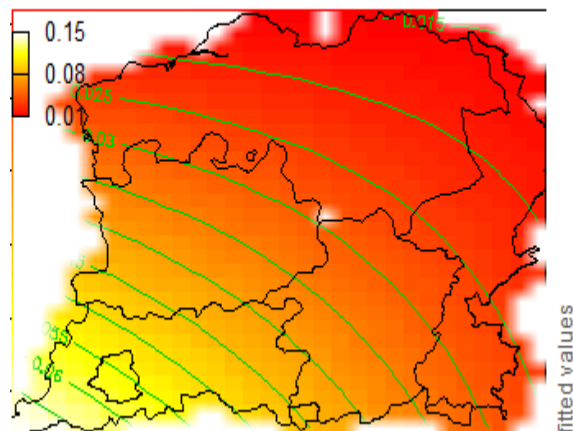
clothing



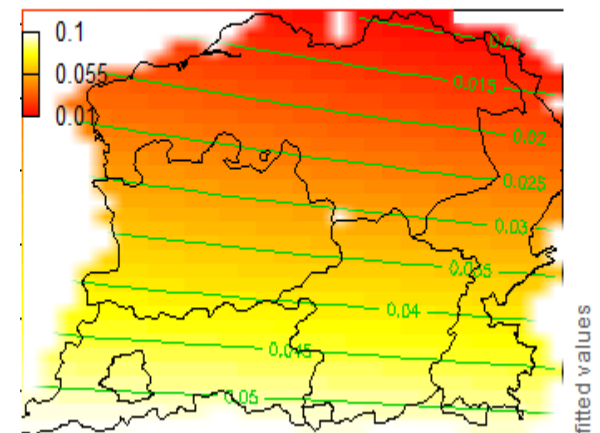
personality & feelings



church & religion



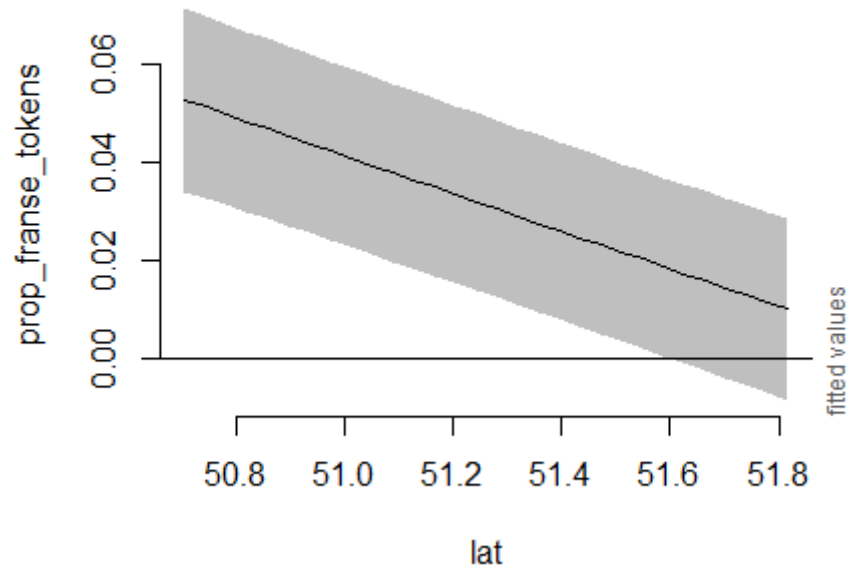
society, school & education



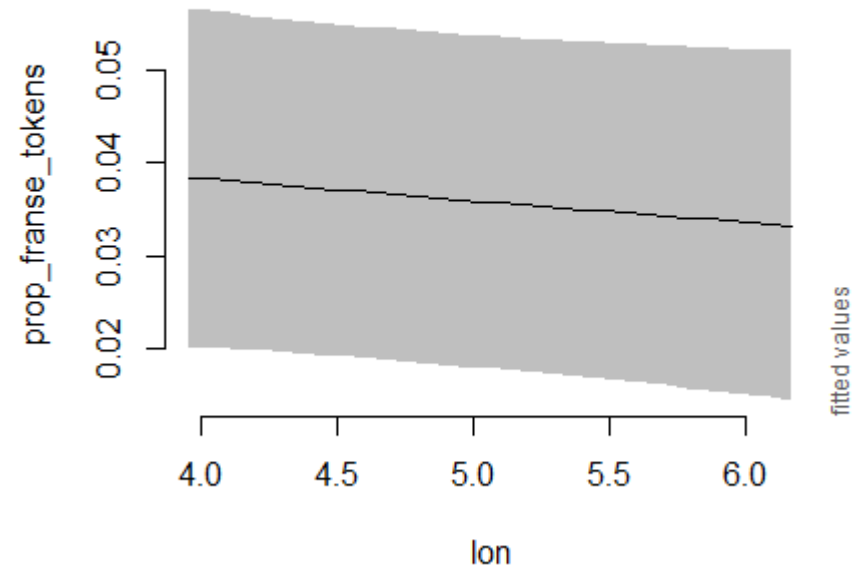
deviance explained: 89.6%

geographical patterns: French

south-north

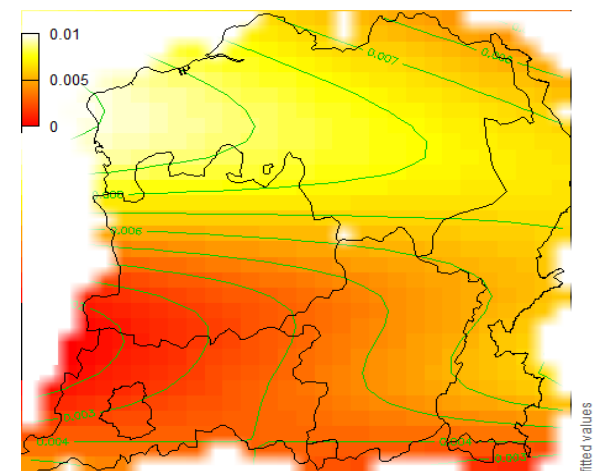
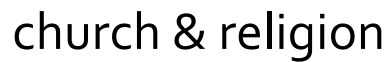


west-east



88% without geography

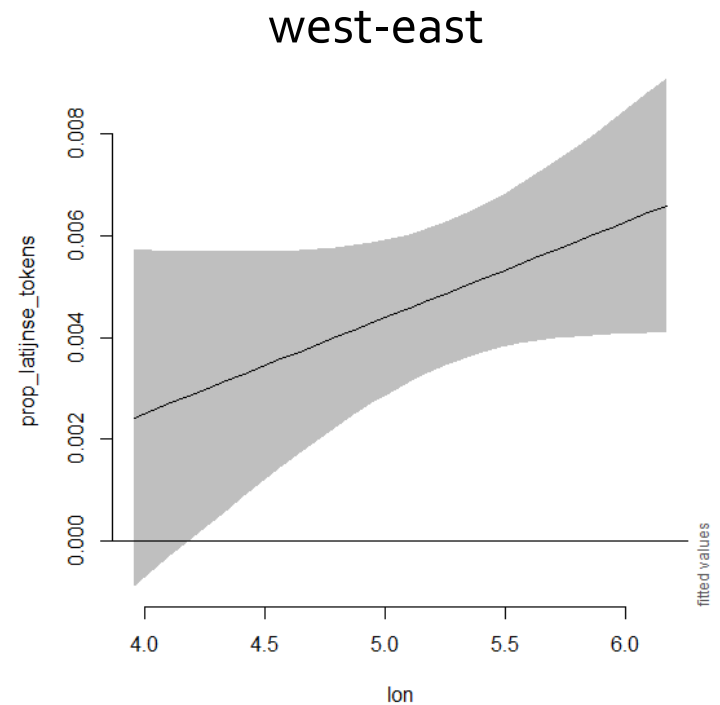
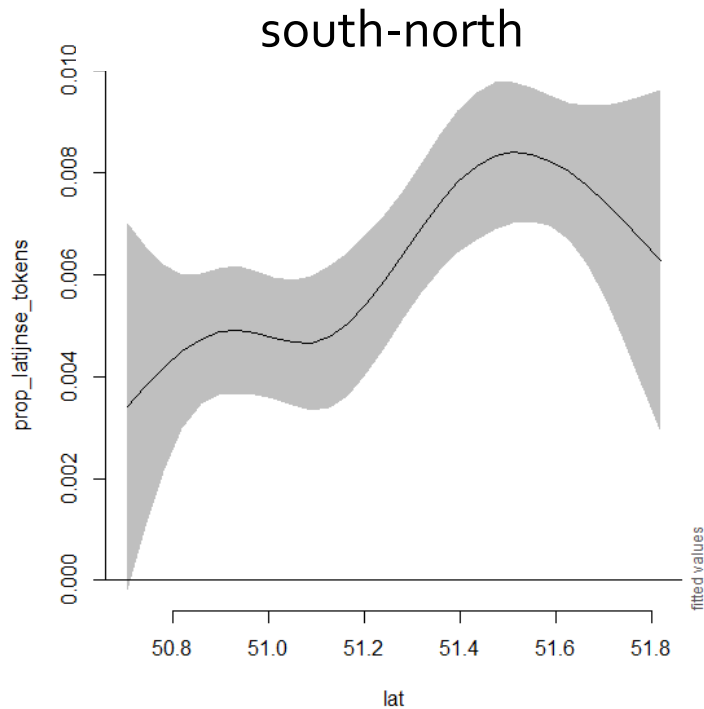
clothing



geographical patterns: Latin

deviance explained : 91.8%

88% without geography

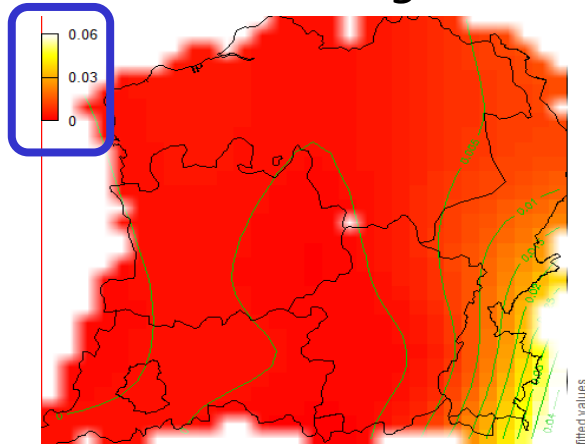


deviance explained : 90.4%

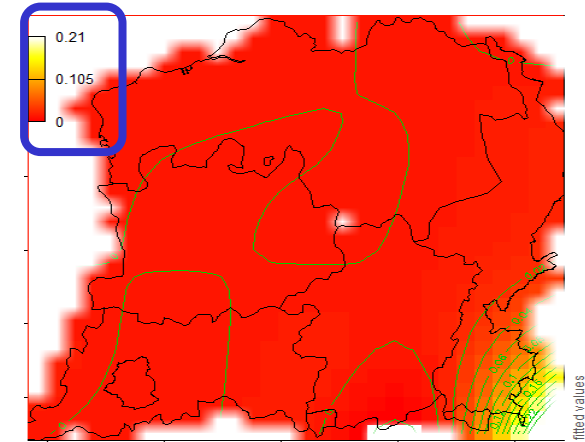
model struggles with general
infrequency of German

semantic patterns: German

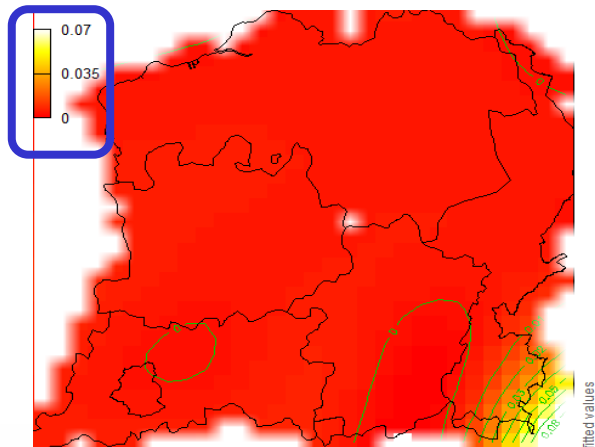
clothing



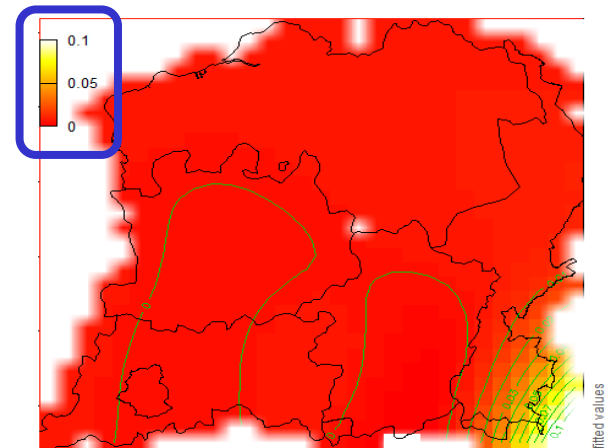
personality & feelings



church & religion



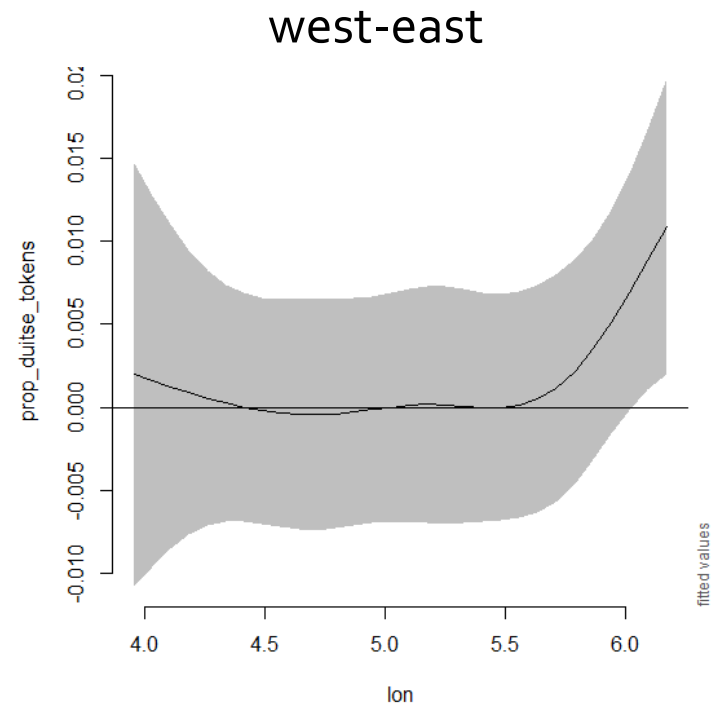
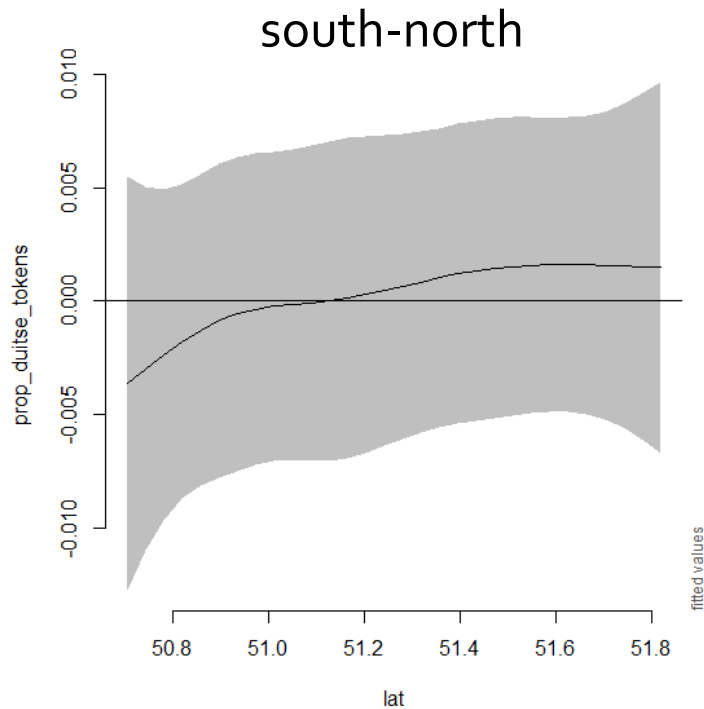
society, school & education



deviance explained : 90.4%

model struggles with general
infrequency of German

geographical patterns: German



conclusions

- naming strategies can help explain geographical heterogeneity in dialects
- expectations partly confirmed:
 - more French in Flanders especially for clothing terminology
 - geography affects the use of Latin, but semantics are more important for borrowings from this source language
 - more German near the German border, but German is only frequently used in a few locations

Merci! Vielen Dank! Gratias vobis ago!

More information:

karlien.franco@kuleuven.be

<http://wwwling.arts.kuleuven.be/qlvl/karlien>

